*THINK INDIA JOURNAL*

ISSN:0971-1260

Vol-22-Issue-41-December-2019

# Object Detection and Identification Using Deep Learning

**Shashanka.P[1], M. Purushotham[2], V. Goutham[3] and K. Ramya Laxmi[4]**
Department of Computer Science and Engineering, Sreyas Institute of Engineering and Technology, Nagole, Hyderabad, Telangana

**ABSTRACT**
*Efficient and correct object detection has been a crucial topic within the advancement of laptop vision systems. With the appearance of deep learning techniques, the accuracy for object detection has inflated drastically. The project aims to include progressive technique for object detection with the goal of achieving high accuracy with a period performance. a significant challenge in mologyany of the article detection systems is that the dependency on different laptop vision techniques for serving to the deep learning primarily based approach, that results in slow and non-optimal performance. during this project, we have a tendency to use a very deep learning primarily based approach to resolve the matter of object detection in Associate in Nursing end-to-end fashion. The network is trained on the foremost difficult publically obtainable dataset , on that a object detection challenge is conducted annually. The ensuing system is quick and correct, therefore aiding those applications that need object detection.*

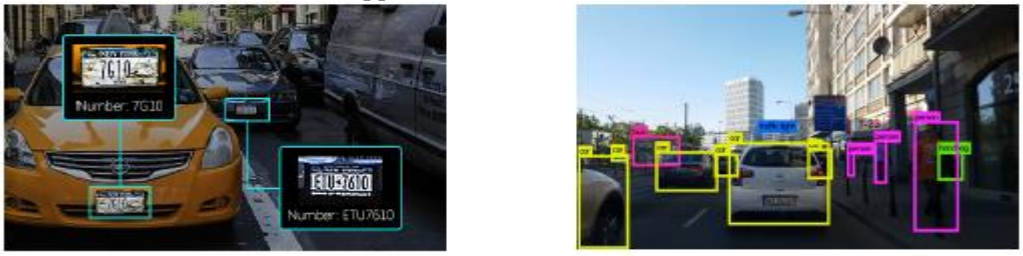
*Keywords: CNN, RCNN, Classification, Regression, SSID design and Deep learning*

## INTRODUCTION

Many issues in laptop vision were saturating on their accuracy before a decade. However, with the increase of deep learning techniques, the accuracy of those issues drastically improved. one amongst the most important downside was that of image classification, that is outlined as predicting the category of the image. a rather difficult downside is that of image localization, wherever the image contains one object and therefore the system ought to predict the category of the situation of the article within the image (a bounding box round the object). The additional difficult downside (this project), of object detection involves each classification and localization. during this case, the input to the system are a image, and therefore the output are a bounding box appreciate all the objects within the image, along side the category of object in every box.

## 1 APPLICATIONS

A acknowledge application of object detection is face detection that's employed in the majority the mobile cameras. A additional generalized (multi-class) application may be employed in autonomous driving wherever a range of objects got to be detected. additionally it's a vital role to play in police investigation systems. These systems may be integrated with different tasks like cause estimation wherever the primary stage within the pipeline is to sight the article, then the second stage are to estimate cause within the detected region. It may be used for chase objects and therefore may be employed in AI and medical applications. therefore this downside serves a mess of applications.




Surveillance Autonomous Vehicles

Fig: application of object detection


Copyright © 2019Authors

ISSN:0971-1260

Vol-22-Issue-41-December-2019

**PROBLEM STATEMENT**

The major challenge during this downside is that of the variable dimension of the output that is caused thanks to the variable variety of objects which will be gift in any given input image. Any general machine learning task needs a set dimension of input and output for the model to be trained. Another necessary obstacle for widespread adoption of object detection systems is that the demand of time period (¿30fps) whereas being correct in detection. The a lot of advanced the model is, the longer it needs for inference; and also the less advanced the model is, the less is that the accuracy. This trade-off between accuracy and performance must be chosen as per the applying. the matter involves classification moreover as regression, leading the model to be learnt at the same time. This adds to the quality of the matter.

**2 CONNECTED WORKS**

There has been a great deal of labor in object detection exploitation ancient laptop vision techniques (sliding windows, deformable half models). However, they lack the accuracy of deep learning primarily based techniques. Among the deep learning primarily based techniques, 2 broad category of ways area unit prevalent: 2 stage detection (RCNN [1], quick RCNN [2], quicker RCNN [3]) and unified detection (Yolo [4], SSD [5]). the most important ideas concerned in these techniques are explained below

**2.1 BOUNDING BOX**

The bounding box could be a parallelogram drawn on the image that tightly fits the thing within the image. A bounding box exists for each instance of each object within the image. For the box, four numbers (center x, center y, width, height) area unit foreseen. this will be trained employing a distance live between foreseen and ground truth bounding box. the gap live could be a jaccard distance that computes intersection over union between the expected and ground truth boxes as shown

Jaccard distance

**2.2 CLASSIFICATION + REGRESSION**

The bounding box is expected pattern regression and so the class within the bounding box is expected pattern classification. this can be done by pattern various algorithms in deep learning.

The regression model provides the basics for classification and identification.

**2.3 TWO-STAGE METHODOLOGY**

In this case, the proposals unit extracted pattern another portable computer vision technique thus resized to mounted input for the classification network, that acts as a feature extractor. Then Associate in Nursing SVM is trained to classify between object and background (one SVM for each class). jointly a bounding box regressor is trained that outputs some some correction (offsets) for proposal boxes. the final set up is These methods unit really correct but unit computationally intensive (low fps)

**2.4 UNIFIED METHODOLOGY**

The distinction here is that instead of producing proposals; pre-define a bunch of boxes to look for objects. pattern convolution feature maps from later layers of the network, run another network over these feature maps to predict class scores and bounding box offsets.

The steps unit mentioned below:

1. Train a CNN with regression and classification objective.
2. Gather activation from later layers to infer classification and placement with a completely connected or convolution layers.
3. throughout work, use jaccard distance to relate predictions with all-time low truth.
4. throughout thinking, use non-maxima suppression to filter multiple boxes around the same object.

**3 APPROACH**

The network used in this project is based on Single shot detection (SSD).



Copyright © 2019Authors

ISSN:0971-1260

Vol-22-Issue-41-December-2019

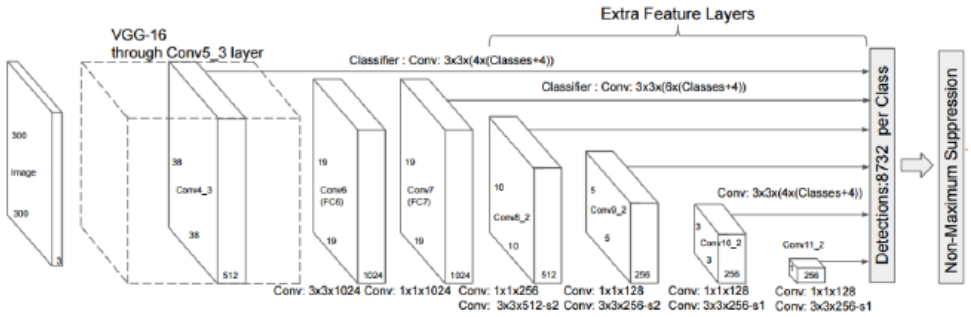



Figure 7: SSD Architecture

The SSD ordinarily starts with a VGG [6] model, that is reborn to a completely convolutional network. Then we have a tendency to attach some additional convolutional layers, that facilitate to handle larger objects. The output at the VGG network may be a 38x38 feature map (conv4 3).

The accessorial layers manufacture 19x19, 10x10, 5x5, 3x3, 1x1 feature maps. of these feature maps ar used for predicting bounding boxes at varied scales (later layers accountable for larger objects). so the plan of SSD. a number of the activations ar passed to the sub-network that acts as a classifier and a localizer.

Anchors (collection of boxes overlaid on image at completely different abstraction locations, scales and facet ratios) act as reference points on ground truth pictures. A model is trained to form 2 predictions for every anchor: • A distinct category • endless offset by that the anchor has to be shifted to suit the ground-truth bounding box

During coaching SSD matches ground truth annotations with anchors. every component of the feature map (cell) features a variety of anchors related to it. Any associate degreechor with an IOU (jaccard distance) larger than zero.5 is taken into account a match. take into account the case, wherever the cat has 2 anchors matched and therefore the dog has one anchor matched. Note that each are matched on completely different feature maps.

The loss perform used is that the multi-box classification and regression loss. The classification loss used is that the softmax cross entropy and, for regression the sleek L1 loss is employed. throughout prediction, non-maxima suppression is employed to filter multiple boxes per object which will be matched.

**Table-1: Pooling layers used for object detection. Pooling layer Description**

| | |
|---|---|
| Max pooling | It is widely used pooling in CNNs. It takes maximum value from the selected image patch and place in the matrix storing the maximum values from other image patches. |
| Average pooling | This pooling averages the neighborhood pixels. |
| Deformation pooling [40] | Deformable pooling has ability to extract deformable properties, geometric constraints of the objects. Spatial pyramid pooling [53] This pooling performs down-sampling of the image and produces feature vector with a fixed length. |
| | This feature vector can be used for object detection without making any deformations on the original image. This pooling is robust to object deformations. |
| Scale dependent pooling [54] | This pooling handles scale variation in object detection and helps to improve the accuracy of detection. |

## 4 EXPERIMENTAL RESULTS

### 4.1 DATASET

For the purpose of this project, the publicly available PASCAL VOC dataset will be used. It consists of 10k annotated images with 20 object classes with 25k object annotations (xml format). These images are



Copyright © 2019Authors

ISSN:0971-1260

Vol-22-Issue-41-December-2019

downloaded from flickr. This dataset is used in the PASCAL VOC Challenge which runs every year since 2006.

### 4.2 IMPLEMENTATION DETAILS

The project is implemented in python 3. Tensorflow was used for training the deep network and OpenCV was used for image pre-processing. The system specifications on which the model is trained and evaluated are mentioned as follows: CPU - Intel Core i7-7700 3.60 GHz, RAM - 32 Gb, GPU - Nvidia Titan Xp.

#### 4.2.1 PRE-PROCESSING

The annotated data is provided in xml format, which is read and stored into a pickle file along with the images so that reading can be faster. Also the images are resized to a fixed size.

#### 4.2.2 NETWORK

The model consists of the bottom webwork derived from VGG net so the changed convolution layers for fine-tuning so the classifier and localizer networks. This creates a deep network that is trained end-to-end on the dataset.

The system handles illumination variations therefore providing a strong detection constant person is standing within the shade so within the sunny atmosphere.

However, occlusion creates a retardant for detection, the occluded birds don't seem to be detected properly. conjointly larger object dominated once gift in conjunction with little objects. this might be the explanation for the common preciseness of smaller objects to be less compared to larger objects. This has been reportable within the next section

### 4.3 MEASURE

The analysis metric used is mean average preciseness (mAP). For a given category, precisionrecall curve is computed. Recall is outlined because the proportion of all positive examples hierarchical on top of a given rank. preciseness is that the proportion of all examples on top of that rank that square measure from the positive category. The AP summarizes the form of the precision-recall curve, and is outlined because the mean preciseness at a collection of 11 equally spaced recall levels [0, 0.1, ... 1].

Thus to get a high score, high preciseness is desired in any respect levels of recall. This live is best than space underneath curve (AUC) as a result of it provides importance to the sensitivity. The detections were allotted to ground truth objects and judged to be true/false positives by activity bounding box overlap. To be thought-about an accurate detection, the world of overlap between the anticipated bounding box and ground truth bounding box should exceed a threshold. The output of the detections allotted to ground truth objects satisfying the overlap criterion were hierarchical so as of (decreasing) confidence output. Multiple detections of constant object in a picture were thought-about false detections, i.e.

5 detections of one object counted as one true positive and four false positives. If no prediction is formed for a picture

then it's thought-about a false negative. the common preciseness for all the article classes square measure reportable in Table three. The mAP for the PASCAL VOC dataset was found to be zero.633. this progressive best mAP worth is reportable to be zero.739.

## CONCLUSION

An correct and economical object detection system has been developed that achieves comparable metrics with the prevailing progressive system. This project uses recent techniques within the field of pc vision and deep learning. Custom dataset was created victimization labelImg and therefore the analysis was consistent. this could be employed in period of time applications that need object detection for pre-processing in their pipeline. a very important scope would be to coach the system on a video sequence for usage in trailing applications. Addition of a temporally consistent network would alter sleek detection and a lot of optimum than per-frame detection.



Copyright © 2019Authors

ISSN:0971-1260

Vol-22-Issue-41-December-2019

**REFERENCES**

1. Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra leader. wealthy feature hierarchies for correct object detection and linguistics segmentation. within the IEEE Conference on laptop Vision and Pattern Recognition (CVPR), 2014.
2. Ross Girshick. Fast R-CNN. In International Conference on laptop Vision (ICCV), 2015. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. quicker R-CNN: Towards realtime object detection with region proposal networks. In Advances in Neural informatics Systems (NIPS), 2015.
3. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. you merely look once: Unified, period of time object detection. within the IEEE Conference on laptop Vision and Pattern Recognition (CVPR), 2016.
4. Wei dynasty Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, ChengYang Fu, and Alexander C. Berg. SSD: Single shot multibox detector. In ECCV, 2016.
5. Tibeto-Burman Simonyan and Apostle Zisserman. terribly deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
6. Ramya laxmi K., Pallavi S., Ramya N. (2020) A Hybrid Approach of Wavelet Transform Using Lifting Scheme and Discrete Wavelet Transform Technique for Image Processing. In: Satapathy S., Raju K., Shyamala K., Krishna D., Favorskaya M. (eds) Advances in Decision Sciences, Image Processing, Security and Computer Vision. ICETE 2019. Learning and Analytics in Intelligent Systems, vol 3. Springer. (Scopus)
7. Sumati Pathak, Rohit Raja, Vaibhav Sharma, and K. Ramya Laxmi, (2019) A Framework Of ICT Implementation On Higher Educational Institution With Data Mining Approach, European Journal of Engineering Research and Science, Vol. 4, Iss. 5, pp. 34-38, Publication date 13/5/2019, ISSN (Online) : 2506-8016.
8. K. Ramya Laxmi, N Ramya, S. Pallavi, (2018) A Survey on Automatically Mining Facets for Queries from their search Results, International Journal of Management Technology and Engineering IJMTE Vol. 8, Iss. 7 July 2018. ISSN NO: 2249-7455 (UGC Approved ).
9. S. Pallavi, K. Ramya Laxmi, N. Ramya, Rohit Raja (2018), Study and Analysis of Modified Mean Shift Method and Kalman Filter for Moving object Detection and Tracking, Published in 3rd International Conference on Computational Intelligence and Informatics (ICCII-2018), held during 28-29 Dec 2018.
10. K. Ramya laxmi, S. Pallavi, N. Ramya, (2019) A Hybrid Approach of Wavelet Transform using Lifting Scheme and Discrete Wavelet Transform Technique for image processing, 2nd National Conference on Cyber Security, Image Processing, Graphics, Mobility and Analytics (NCCSIGMA 2019), Organized by Department of CSE at CMR Technical Campus, Hyderabad in association with DIV – 5 Education & Research, CSI India from 24th – 25th Jan 2019.
11. K. Ramya laxmi, N. Ramya, S. Pallavi, K. Madhuravani, (2019) Study and Analysis of Apriori and K-Means Algorithms for Web Mining, 8th International Conference On "Innovations In Electronics & Communication Engineering (ICIECE-2019)" On August 02-03, 2019.
12. K. Ramya laxmi, Marri Abhinandhan Reddy, CH. Shivasai, P. SandeepReddy, 8th International Conference On "Innovations In Electronics & Communication Engineering (ICIECE-2019)" On August 02-03, 2019.



Copyright © 2019Authors